

Проект типологической базы данных уральских языков

Ф. З. Хаваш (F. Z. Havas),
профессор лингвистики,
заведующий кафедрой финно-угроведения
Университета им. Этвёша Лоранда
(г. Будапешт, Венгрия)

Изначально уралистика предполагала генеалогическое сравнение языков, в принципе такое положение сохраняется и в наши дни. Это понятно, поскольку ее исходная задача состояла в открытии уральских языков для науки, в доказательстве их родства, в конструировании родословного древа, реконструкции языка-основы, одним словом, уралистика относилась к сравнительно-историческому языкознанию. Остальные два компаративных подхода, ареальный и типологический, в прошлом не отделялись от генеалогического. Хотя они не пользовались таким же вниманием, все же нельзя отрицать, что в данном отношении в последние десятилетия произошли некоторые изменения в лучшую сторону. Задача уралистики как языковедческой науки состоит в конечном счете в научном рассмотрении уральских языков и их презентации в ряду языков мира. Однако выполнять указанную задачу сегодня уже невозможно, если не выходить за рамки уральских языков. Тот, кто желает приобрести углубленные знания в области уралистики, должен выйти за рамки внутреннего исследования этой языковой семьи, должен смотреть дальше во времени и пространстве. Желаемый обширный горизонт может быть достигнут именно путем ареального и типологического исследования уральских языков. Настоящий проект обращен на их типологическую разработку.

Типология уральских языков решает, как правило, две задачи. Одна из них заключается в типологиче-

ском рассмотрении этих языков для определения их места среди языков мира. Вторая состоит в том, чтобы типологически представить уральские языки для общей лингвистической типологии, или, иными словами, приблизить друг к другу общую типологию языков как таковую и уралистику. Сегодня мы сталкиваемся с очевидными недостатками в обеих отраслях. При описании уральских языков, а также объяснении их диахронных связей часто недостает учета типологических знаний, касающихся других языков мира; в то же время в типологических характеристиках уральских языков, присутствующих в анализах языков мира, написанных не уралистами, специалисты часто находят странные, поверхностные или же вовсе неправильные высказывания.

Представляется логичным преодоление выявленных недостатков путем обоюдного сближения дисциплин. Решение этой задачи в большей мере должно приходиться на долю уралистов, ориентирующихся на типологические вопросы (ввиду их непосредственной заинтересованности), чем на типологов, ориентирующихся на уралистику.

Для типологического описания уральской языковой семьи видится ряд возможностей. Можно, например, иметь целью составление полного типологического описания каждого уральского языка на основе совсем нового анализа, конечно, следуя какой-то заранее определенной единой концепции. Такое представление не является выдумкой, подобные попытки в прошлом предпринимались. Стоит привести хотя бы самый известный пример: Бернард Комри и Норвал Смит в 1977 г. предложили подробный вопросник, позволяющий получить объемные характеристики отдельных языков где-то на полпути между дескриптивной грамматикой и типологической характеристикой. (Заметим, что такие работы были выполнены по венгерскому и финскому языкам. Это интересные исследования, но в силу «жанровой неопределенности» они могут

Исходная задача уралистики состояла в открытии уральских языков для науки, в доказательстве их родства, в конструировании родословного древа, реконструкции языка-основы.

быть использованы, скорее, только как пособия для типологической характеристики.)

Задача могла быть решена и другим путем, если бы в свое время проект Майтинской и Феенкера, предполагавший создание серии монографий с единообразным описанием уральских диалектов, был претворен в жизнь. Мы имеем в виду, что в идеальном – и потому никогда не осуществимом – случае работа типолога начинается там, где работа дескриптивиста заканчивается. Если бы мы располагали этим материалом, типологическая «выписка» из него была бы осуществима сравнительно легко. Впрочем, в виде известного типа справочников такие описания с единой структурой выходили в свет, а именно в монографиях и сериях советской, а потом Российской академии наук (таких как «Языки народов СССР» и «Языки мира»). Несмотря на все это, мы так и не располагаем достаточно подробным и по своей структуре единым образом организованным описанием уральских языков, из которого мог бы исходить типологический анализ.

Третий путь решения задачи, который может быть поставлен на повестку дня в уралистике, – проект создания типологической базы данных уральских языков. Я выступил с инициативой разработки такого проекта, и сегодня можно сказать, что для его осуществления уже сформировалась высокопрофессиональная международная группа.

Идея создания типологической базы данных возникла у меня в начале этого десятилетия, была обдумана и в 2005 г. предложена уралистам на X Международном конгрессе финно-угроведов в Йошкар-Оле. Собралось около тридцати коллег, они выслушали проект с интересом, некоторые даже с энтузиазмом, многие записались в список возможных участников. Большое значение имело создание собственного веб-сайта проекта (<http://www.univie.ac.at/urtypol/index.html>), который продолжает работать и ныне, а в будущем, по моим предположениям, его роль еще более возрастет.

Вначале предполагалось создать базу данных за пять лет. Однако это предположение оказалось утопическим. Хотя среди замечаний, приходящих на веб-сайт, было много положительных идей, в период после конгресса работа так и не началась. Стало ясно, что необходим новый импульс. Именно поэтому была выдвинута идея организации так называемой стартовой конференции, которая состоялась в Вене

осенью 2008 г. Но прежде чем рассказать о данном мероприятии, я хочу кратко представить суть идеи самой типологической базы данных уральских языков, как это было изложено мной во вступительном докладе на конференции (полный текст на английском языке см. на вышеупомянутом веб-сайте).

Типологическую базу данных при первом знакомстве можно представить как воображаемую онлайн-таблицу, т. е. множество данных, размещенных в столбцах и строках. Столбцы таблицы – уральские (финно-угорские и самодийские) языки: венгерский, марийский, нганасанский и т. д. Строки таблицы содержат типологические параметры, например, основной порядок слов, притяжательная конструкция, условные придаточные предложения и т. д. Количество столбцов – минимум 18, столько, сколько насчитывается уральских языков. Некоторые языки можно разделить на несколько диалектов, в силу чего число столбцов увеличится. Количество же строк не ограничено; если мы придем к полному составу типологических параметров языков, их может насчитываться несколько сотен.

Таким образом, таблица состоит из клеток, каждая из которых находится в определенной точке пересечения строки и столбца, как в любой таблице. Каждая клетка несет в себе значение типологического параметра (критерия) соответствующей строки. Если, например, 48-я строка представляет тип синтагм с числительными, а 7-й столбец принадлежит

Проект создания типологической базы данных уральских языков

- идея предложена в 2005 году на 10-м Международном конгрессе финно-угроведов в Йошкар-Оле
- создание собственного веб-сайта проекта
<http://www.univie.ac.at/urtypol/index.html>
- конференция-запуск: 26–27 сентября 2008, Вена
<http://www.univie.ac.at/urtypol/conf.html>

- ИЗ ВЫВОДОВ:

- запустить пилотные проекты по подгруппам уральской языковой семьи
- представить опыт и имеющиеся материалы на 11 Международном конгрессе финно-угроведов в Пилишчабе (2010 г.)



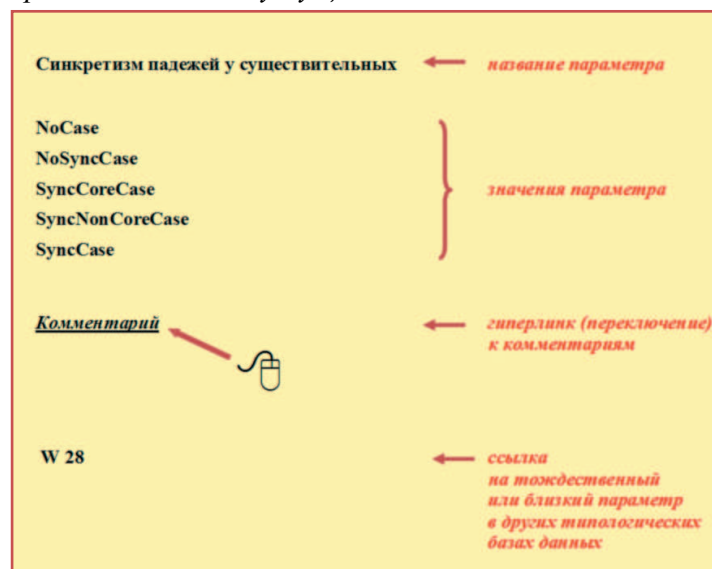
Югорский писатель Ю. Шесталов и профессор Ф. Хаваш (справа).
IV Всероссийская конференция финно-угроведов «Языки и культура
финно-угорских народов в условиях глобализации».
Ханты-Мансийск, 2009 г.

эзя-мордовскому языку, то клетка, находящаяся в точке их пересечения, т. е. в 48-й строке 7-го столбца, содержит типологическую характеристику синтагмы с числительными эзя-мордовского языка. В клетках, находящихся непосредственно слева и справа от нее, т. е. в той же строке, можно найти синтагмы с числительными марийского и мокша-мордовского языков, а в клетках непосредственно под и над ней – другие типологические параметры, например, притяжательные или адпозиционные синтагмы в эзя-мордовском языке. Можно посмотреть не только одну строку один столбец таблицы, но сразу несколько строк и столбцов, если в этом возникнет необходимость. В то же время из них можно выделить и любой частный состав. Например, если мы интересуемся лишь одной строкой, пусть это снова будут конструкции с числительными, но в прибалтийско-финских языках, то таблица будет показывать только данные, соответствующие столбцам прибалтийско-финских языков. Таким же образом мы можем выделить два столбца, находящихся на любом расстоянии друг от друга в таблице, например, типологические характеристики конструкций с числительными в финском и эзя-мордовском языке, пропустив все остальные уральские языки. Выделение относится, конечно, и к сокращению анализируемой части столбцов. Мы можем выделить описание всех определительных синтагм того или иного языка (от 5 до 10 клеток из данного столбца), при этом не обращая внимания на другие данные столбца. Фильтрация обоих видов может применяться и одновременно, если мы хотим выделить больше одного, но все-таки ограниченное количество типологических параметров больше чем

в одном, но все-таки в ограниченном количестве языков, например, все определительные синтагмы в пермских языках.

Щелкнув на любую из клеток первого, крайнего левого, столбца, т. е. типологических параметров, найдем общее описание сути данного типологического параметра и его возможных вариантов в языках мира. Там же логично было бы привести основную литературу по вопросу, а также ссылки на базы данных, уже представленных в Интернете. Это было бы важно для специалистов-уралистов, имеющих умеренные познания по общим типологическим сведениям, относящимся к данному языковому явлению. Описание и/или литературу и ссылки можно приобщить и к столбцам верхней строки, т. е. к отдельным языкам и диалектам, – в интересах типологов – не уралистов, желающих побольше знать о тех языках, из которых они и берут свои примеры или проверяют данные других авторов.

Целесообразно вкратце рассмотреть оформление параметров. Сделаем это на примере параметра *Синкретизм падежей у существительных*.



В первой строке задано название параметра. Далее приводятся возможные значения параметра, причем предполагается, что из них одно и только одно приписывается каждому языку мира, в нашем случае исследуемому финно-угорскому языку (диалекту). Значения заданы в кодировке, основанной на англоязычных определениях. Ниже следует место переключения к комментариям, а затем ссылка на тождественный или близкий параметр в других типологических базах данных, в частности на статью 28 WALS – Всемирного атласа лингвистических структур (вместе с типологической базой данных Драйера он послужил источником для подбора состава и, отчасти, оформления параметров).

Щелкнув на гиперссылку *Комментарий*, мы можем увидеть объяснение к параметру и его значениям.

Оно может быть оформлено примерно следующим образом:

Синкретизм падежей – это их парадигматический синтетизм. Синкретизм падежей имен существительных наблюдается, когда в системе склонения одно и то же падежное оформление (напр. тот же самый падежный аффикс) выполняет функцию нескольких падежей. Наличие какого-то падежа (падежной функции) устанавливается в языке, если в последнем имеет место склонение существительных и данная функция приобретает оформление, отличное от других падежей, хотя бы у некоторых групп существительных или их подклассов.

Типы падежного синкретизма:

NoCase:

в языке нет склонения существительных

NoSyncCase:

в языке есть склонение существительных и падежи
никогда не синкретичны

SyncCoreCase:

в языке есть склонение существительных и
синкретизм имеет место исключительно у основных
падежей

SyncNonCoreCase:

в языке есть склонение существительных и
синкретизм имеет место исключительно у неосновных
падежей

SyncCase:

в языке есть склонение существительных и синкретизм имеет место как у основных, так и у неосновных падежей

Комментарии, таким образом, выполняют тройную функцию. Во-первых, они объясняют суть самого параметра. Во-вторых, осведомляют о том, как раскрыть данное явление в языке (указывая, если необходимо, на возможные ошибки идентификации явления, например предупреждая о подобных, но к данному параметру не относящихся явлениях). В-третьих, объясняют сами значения, т. е. указывают, в каком случае данному языку то или иное значение параметра приписывается.

Перейдя затем к клеткам самой таблицы, можно убедиться, что в результате их заполнения в них появляется как раз то значение параметра, которое установлено исследователем для данного языка (диалекта), т. е. подходящий код. (Например, в русском языке для рассматриваемого параметра это было бы SyncCoreCase, ввиду того что такое совпадение падежных форм в нем имеется, но наблюдается только у именительного и винительного падежей некоторых

типов существительных.) Пояснение значений параметров в кодировке необходимо для того, чтобы наша база данных могла быть связана с уже существующими базами данных (и обеспечивался «свободный переход» в оба направления). В то же время, чтобы база была доступна для широкого круга лингвистов, т. е. не только для типологов, ее содержание должно быть представлено и в текстуальной форме, с подробными объяснениями и набором примеров. Такой текст и появляется, если щелкнуть на код в клетке. Объяснительный текст представляет «фоновую» информацию к данному параметру конкретного языка (подробное описание характеризуемого явления с анализом отдельных примеров, указание на вопросы, нуждающиеся в дальнейших исследованиях, история заполнения клетки (авторство), важнейшая литература по вопросу). Его можно составить сначала на русском языке, впоследствии он будет переведен на английский. Вообще, я представляю нашу базу данных работающей на двух языках (за исключением самой кодировки).

Таблица может находиться в процессе постоянного становления (потому, кстати, она и должна существовать в электронном виде, онлайн). Например, если кто-то предложит включение нового параметра, то в таблице возникнет новая строка, вначале с пустыми клетками. Гораздо менее вероятно предложение нового столбца (диалекта), но и это не вызовет проблем, получится только серия новых пустых клеток, ждущих заполнения. Таблица непрерывно пополняется и количественно, и качественно, в то же самое время позволяя бесперебойно пользоваться собой на основе уже готового материала. Итак, типологическая база данных уральских языков может составляться поколениями лингвистов, которые одновременно могут и пользоваться ею.

Вот в общих чертах идея, которую я представил на упомянутой Венской конференции. Тем, кто интересуется подробностями, предлагаю изучить вышеуказанный веб-сайт, где можно найти не только имена докладчиков и подробную программу конференции, но и раздаточные материалы и слайды все накапливающегося количества докладов. Ограничусь здесь только описанием своеобразной концепции мероприятия. Это была конференция в виде семинара (workshop). Поскольку она представляла собой первый шаг на пути к созданию *типологической* базы данных *уральских* языков, участниками конференции были специалисты отчасти общей лингвистической типологии, отчасти уральских языков, причем каждый из них был специально приглашен. Более того, были заранее определены не только докладчики, но и темы их докладов. Типологи (в первую очередь один из самых известных – Метью Драйер) сообщали о принципах и практике существующих типологических баз данных, уралисты же выступали с докладами общего,

вводного характера, касающимися типологического описания уральских языков. Затем – и в этом состояла настоящая новизна – были представлены спаренные доклады. Суть названного «жанра» состояла в том, что по определенной тематике было прочитано по два доклада: типолог говорил о том, какими основными знаниями располагает лингвистическая типология в области данного явления вообще в языках мира, а потом уралист давал отчет о том, как и в каких разновидностях наблюдается то же самое явление в уральских языках. Цель такой системы работы заключалась в представлении, сопоставлении и, если возможно, приближении друг к другу двух подходов, типолога и уралиста, ведь разработать типологическую базу данных уральских языков можно только гармонизацией, интеграцией обоих подходов.

Задача состоит в том, чтобы в предстоящий период запустить в какой-то мере гармоничные, но друг от друга независимые, предварительные или пробные (по-английской традиции они называются пилотными), «малые» проекты, которые, по моим представлениям, можно будет включить в общий «большой» проект. Любая такая работа требует сотрудничества типологов и специалистов уральских языков. (Сводя дело к самой сути, типолог спрашивает, а уралист отвечает. Конечно, может получиться, что, с точки зрения уралиста, вопрос типолога поставлен неправильно или он не ставит нужный вопрос, в таком случае возникает ситуация обратной связи, когда уралистика обогащает типологический подход, модифицируя и пополняя сами типологические параметры.)

Здесь открывается поле для пилотных проектов. Мы можем выделить базу данных или ее тематическую часть в качестве образца (например, параметры/вопросы, касающиеся морфологии или определенных подобластей синтаксиса), определить круг исследуемых языков (или ограничиться лишь одним уральским языком для рассмотрения с данной позиции). С формальной стороны такой проект может быть осуществлен в рамках тематических мини-конференций и/или тематических томов, при условии, что в итоге результаты можно будет свести в единый документ (мини-базу данных, таблицу).

Мое конкретное предложение сводится к тому, чтобы запустить пилотные проекты в области разных ветвей уральской языковой семьи. О запуске такого пилотного проекта, направленного на создание типологической базы данных для угорских языков, в принципе мы уже договорились с коллегами из Югорского государственного университета, но такие же начинания были бы желательны и в отношении пермских (удмуртского, коми, коми-пермяцкого), волжских (обоих мордовских и марийского), а также самодийских языков. Разработку каждой группы целесообразно вести в рамках и под управлением компетентных университетов (их компетентных кафедр) на местах, например, Удмурт-

ского государственного университета в Ижевске, Марийского государственного университета в Йошкар-Оле и т. д., или же в соответствующих институтах языкознания. Окончательной целью проектов было бы составление типологической базы данных групп языков, но сферу деятельности в начальный период можно ограничить морфологией или/и определенными областями синтаксиса. Разработка параметров и предоставление их в распоряжение местных исследователей – задача моя и моих сотрудников. Над этой задачей мы трудимся уже много месяцев, основной состав параметров предполагается представить в 2010 г. Не буду скрывать, что при работе над параметрами возникло затруднительное обстоятельство: при первых попытках перевода комментариев на русский язык мы столкнулись с проблемой расхождения лингвистических традиций. Комментарии, составленные с использованием распространенной в мировой специализированной литературе типологической терминологии (в том числе употребляемой в моих главных источниках – в WALS и Драйером), должны сначала подвергаться своего рода «русификации», чтобы они могли стать вполне освоенными и употребительными для финно-угроведов, работающих, как правило, на основе традиций российского языкознания. (При этом, конечно, и сами параметры, и их значения должны остаться в оригинальном виде, переработке здесь подвергаются сами комментарии, т. е. объяснения к параметрам и их возможным значениям.) Для преодоления проблемы нужно найти (к тому же добровольного) специалиста, одинаково сведущего в обеих традициях. С его помощью будет достигнуто то оформление комментариев к параметрам, которое может лечь в основу исследовательской работы на местах.

Надеюсь, что проект создания типологической базы данных для уральской языковой семьи в целом и прилагаемых к нему пилотных проектов в частности вызовет интерес со стороны российских коллег – специалистов финно-угорских языков. Без сотрудничества проект неосуществим, а с их помощью мы можем создать такую прочную базу, благодаря которой сокровища наших языков обогатят данные всемирной типологии и общего языкознания, данные о человеческом языке как таковом. При этом наши языки будут изучены и представлены для всех, в том числе для нас самих, на уровне языкознания начала XXI столетия.

Ключевые слова / keywords:

*уральские языки; типология;
база данных*

Ural languages; typology; database

Поступила 19.11.09